Article

Link Prediction Based on Heterogeneous Social Intimacy and Its Application in Social Influencer Integrated Marketing

Shugang Li [1], He Zhu [1], Zhifang Wen [1], Jiayi Li [2], Yuning Zang [1], Jiayi Zhang [1], Ziqian Yan [1] and Yanfang Wei [1,*]

1 School of Management, Shanghai University, Shanghai 200444, China; luck_li@shu.edu.cn (S.L.)
2 Songjiang No. 2 Middle School, Shanghai 201600, China
* Correspondence: weiyanfang@shu.edu.cn

Abstract: The social influencer integrated marketing strategy, which builds social influencers through potential users, has gained widespread attention in the industry. Traditional Scoring Link Prediction Algorithms (SLPA) mainly rely on homogeneous network indicators to predict friend relationships, which cannot provide accurate link prediction results in cold-start situations. To overcome these limitations, the Closeness Heterogeneous Link Prediction Algorithm (CHLPA) is proposed, which uses node closeness centrality to describe the social intimacy of nodes and provides a heterogeneous measure of a network based on this. Three types of heterogeneous indicators of social intimacy were proposed based on the principle of three-degree influence. Due to scarce overlapping node sample data, CHLPA uses gradient boosting trees to select the most suitable index, the second most suitable index, and the third most suitable index from Social Intimacy Heterogeneous Indexes (SIHIs) and SLPAs. Then, these indicators are weighted and combined to predict the likelihood of other node users in the two product circles in an online brand community becoming friends with overlapping node users. Finally, a hill-climbing algorithm is designed based on this to build integrated marketing social influencers, and the effectiveness and robustness of the algorithm are validated.

Keywords: CHLPA; SIHI; integrated marketing; social influencer; gradient boosting trees

MSC: 05C82; 05C90

1. Introduction

In the past decade, online social networks have firmly rooted themselves in our lives, becoming an integral part of it. Almost everyone has a personal profile on some social network. Social networks provide a detailed record of human communication patterns, offer a convenient way to disseminate information, and explore the structure of social networks. Users within social networks form communities or groups [1] based on shared interests, hobbies, professions, locations, or careers, creating what is often referred to as a circle. Companies can leverage these circles to establish corresponding online brand communities [2]. Users within these communities form connections through communication and interaction and, in the case of brand communities, integrating marketing social influencers is defined as a multi-circle overlap node user who can become friends with many people in the product circle and influence the purchase decisions of ordinary node users through their behavior [3,4]. A large amount of research has shown that integrating marketing social influencers has a significant impact on consumers' willingness to join brand pages [5], share electronic word-of-mouth information [6], brand attitudes [7], product evaluations, purchase likelihood [8], and actual purchase behavior [9] through their friend relationships.

Users in a brand community are typically consumers of the same brand. By having their own online community, brands can effectively gather consumers together and increase their brand loyalty. The existence of online communication platforms means that online interactions between different brand communities not only affect the relationships between

consumers, but also between consumers and brands [10]. Through these platforms, user interactions continue to strengthen, and the opinions and recommendations of friends play an increasingly important role in influencing the purchasing decisions of other users and their attitudes towards companies. Therefore, brand communities are crucial for online marketers as they can often encourage consumers to make purchase decisions [11]. In the era of fragmented marketing, companies are still increasingly using highly influential friends, i.e., integrated marketing social influencers, for marketing purposes. Social influencer integrated marketing plays an irreplaceable role in enhancing a company's brand image and achieving a brand's long-term development goals.

Integrated marketing is a strategy that combines the sales of two different branded products by integrating their respective customer bases and conducting joint marketing activities. This strategy can break down the barriers between product circles and significantly improve marketing efficiency, just like the marketing case of beer and diapers, where their correlation can promote each other's sales and achieve the goal of simultaneously increasing sales volume. Nowadays, integrated marketing has attracted widespread attention in the industry, and "everything can be crossed, everything can be connected" has become a common trend. Recently, joint products such as the "Uniqlo $\times$ KAWS" T-shirt and the "NetEase Cloud Music $\times$ Sanqiang" underwear caused a frenzy among the public and set off a storm on social media. "Crossing borders and collaborating with others" has become the slogan of many brands. From real estate and cars to shoes, handbags, lipsticks, and drinks, as long as they are labeled with a joint name, they can always create a wave of excitement.

Using popular social influencers for integrated marketing can indeed achieve the goal of integrated marketing, but the high cost of using these influencers makes it difficult for new brands in their early stages with limited funds. Therefore, this study proposes to build integrated marketing social influencers by cultivating users with potential (i.e., users who currently have few friends but will have many in the future) and recommending them to other users as friends. In social network brand communities, integrated marketing social influencers generally cross different brand circles and are overlapping nodes in different circles [12]. Integrated marketing in brand communities on social networks requires predicting the likelihood of overlapping node users in two circles establishing friendships with other node users in the two circles. If the likelihood is high, then they are recommended to become friends. Because overlapping node users have network characteristics different from other node users in the circle, such as being located in two product circles at the same time and being rare in number, recommending them to establish friendships with other node users will inevitably face the cold-start problem. Existing friend recommendation algorithms for establishing friendships in social networks mainly use the Scoring Link Prediction Algorithm (SLPA), which predicts the likelihood of a link between user nodes based on the similarity of the social network's topological structure, such as the characteristics of common neighbors. However, these SLPA algorithms mainly rely on homogeneous network indicators to predict friendship relationships, do not consider the differences in social intimacy between node pairs that have already established friendships, have low information content, and cannot adaptively generate combination link prediction algorithms based on specific circle structure features, which cannot provide accurate link prediction results in cold-start situations.

To address the challenges posed by the cold-start problem in predicting node-pair friendships with insufficient information, this study proposes a novel approach that leverages heterogeneous indicators to capture the macro structure of the network. Traditional solutions have typically relied on local information only, which can limit the accuracy and reliability of predictions. By utilizing heterogeneous indicators, this study aims to extract hidden information from deep mining network structures and provide a more effective means for predicting node pair friendships while also addressing overlapping node relationships. By considering the macro structure of the network, the proposed approach can increase the amount of reliable information available and help to avoid the cold-start

problem caused by a lack of data. The resulting predictions can be more accurate, enabling better decision-making in various applications.

To characterize social intimacy, this study utilizes the global feature indicator of closeness centrality and proposes the Closeness Heterogeneous Link Prediction Algorithm (CHLPA) to predict the likelihood of a target user becoming friends with other node users. CHLPA uses the variable of node closeness centrality to describe the social intimacy of nodes and provides a heterogeneous measurement of the network based on this. Based on the three-degree influence principle, three types of heterogeneous indicators of social intimacy are proposed, including the self-Social Intimacy Heterogeneous Index (SIHI), common neighbor-based SIHI, and community neighbor-based SIHI.

To address the issue of traditional mining methods being ineffective due to scarce data on overlapping node samples, CHLPA employs the Gradient Boost Decision Tree (GBDT) method. Using the density of nodes and their centrality in local networks, the most suitable index, the second most suitable index, and the third most suitable index from SIHIs and SLPAs are selected. To improve the model's performance and avoid the risk of algorithmic accuracy reduction resulting from blind SIHI combinations, CHLPA applies the GBDT method to obtain a final fit probability by assigning weights to the selected SIHIs and combining them. Composite indexes are used to predict the likelihood of other node users in two product circles in the online brand community becoming friends with overlapping node users.

Finally, this study presents a hill-climbing algorithm for building integrated marketing social influencers, which utilizes CHLPA to predict the likelihood of overlapping node users linking with ordinary node users within two product circles. Based on whether their closeness centrality is improved, the algorithm determines whether to recommend nodes with link potential as friends, thereby achieving the construction of social influencers for integrated marketing. The effectiveness and robustness of the algorithm were verified through datasets from social networks.

In summary, this study presents several key innovations: (1) The application of CHLPA in integrated marketing is novel. (2) The innovative SIHI indicators. (3) The use of the GBDT model to innovatively select and combine indicators. (4) The adaptive selection of indicators based on network characteristics. These innovations have successfully realized the construction of integrated marketing social influencers in online brand communities, providing new ideas and insights for research on social network analysis and marketing strategies.

The remaining parts of this study are arranged as follows: Section 2 presents the link prediction research status; Sections 3 and 4 introduce the CHLPA and the hill-climbing algorithm for constructing "integrated marketing social influencers"; Section 5 verifies the effectiveness of the CHLPA and the hill-climbing algorithm through datasets in social networks; and Section 6 provides a conclusion of this study.

2. Link Prediction Research Status

2.1. Link Prediction

In studies related to link prediction, the scoring method based on common neighbors is the most widely used algorithm for link prediction. Its fundamental principle is to judge the likelihood of a link between nodes based on the high or low scores calculated from common neighbor-related indicators. The higher the score, the greater the likelihood of a link appearing in the future network. Lü and Zhou [13] classified link prediction algorithms into three categories: similarity-based link prediction, maximum likelihood estimation-based link prediction, and probability model-based link prediction. Among these existing link prediction algorithms, the similarity-based algorithm is the most widely used framework for predicting non-existent links, where the score between two nodes is directly defined as their similarity [14].

2.2. Combined Link Prediction

With the development of research, more and more scholars have found that a single link prediction algorithm cannot meet the prediction needs of extremely sparse networks. Therefore, in addition to the common link prediction algorithms mentioned above, many scholars have proposed combined link prediction algorithms to deal with the link prediction problem in sparse networks. For example, He and Liu [15] found that the stability of link prediction algorithms based on local information was usually very low. Therefore, they proposed a link prediction ensemble algorithm (LPEA) based on the Ordered Weighted Averaging (OWA) operator. LPEA used three different OWA operators to assign integration weights to various link prediction algorithms. Wang, Ma [16] and others proposed a parameter-adjustable link prediction algorithm based on community information (CI) and applied it to large-scale complex networks to obtain community information for link prediction. Xiao and Li [17] generalized the LDA topic model and Hidden Naive Bayes algorithm and proposed a three-level hidden Bayesian link prediction (3-HBP) model for link prediction. Also, Bütün and Kaya [18] analyzed user behavior and user relationships to mine users' interest distribution and predicted missing links between users. They proposed a pattern-based supervised link prediction method to improve the link prediction accuracy of the Triad Closeness (TC) measure in directed complex networks. The proposed pattern-based link prediction measure was compared with the TC measure and the latest link prediction measure, and the results confirmed the effectiveness of their proposed measure.

2.3. Heterogeneous Link Prediction

Unlike the homogeneous link prediction algorithms described earlier, in recent years, some scholars have proposed heterogeneous link prediction algorithms to extract deeper information from the network. For example, Zhan and Zhang [19] studied finding influential users across multiple heterogeneous social networks in different sub-communities and proposed a new network diffusion model called "Cross-Network Information Diffusion" (CONFORM). Based on this, they extracted and fused various diffusion links in a heterogeneous network and calculated the probability of user activation. Yudhoatmojo and Budi [20] used edge-weighted and centrality-based weighting to retrieve opinion leaders related to rumor propagation. They examined the importance of each defined edge type in finding opinion leaders through edge weighting and found the weights that can provide more accurate opinion leaders through centrality weighting. Dai and Shang [21] developed a new framework to address the embedding problem in heterogeneous networks by proposing a path-guided random walk strategy related to edges to guide walking between different layers separated by edge types, and then used a heterogeneous jump model to calculate the overall node embedding. The results of quantitative experiments on four public datasets (Amazon, Youtube, DBLP, and Movielens) showed that their method achieved significant improvements in link prediction tasks. Wen [22] designed a link prediction algorithm that combined node structure and sentiment attributes, analyzing the linkages and tweets of some hot topics in the 2014 World Cup on Tencent Weibo and the emotional distribution of each topic's audience. The experimental results showed that the number of users with the same emotion and the emotional distribution of the audience can affect the links between users. Mohdeb and Boubetra [23] proposed a weighted meta-path-based link prediction method, WMPLP, which significantly improved the effect of link prediction by maximizing the information richness in heterogeneous social networks. Cao and Kong [24] proposed an iterative framework for heterogeneous collective link prediction, called HCLP, which utilized diversified and complex link information in heterogeneous homogeneous information networks to predict multiple types of links. Empirical studies based on real tasks have shown that this method can effectively improve the link prediction performance in heterogeneous information networks.

The link prediction algorithm is a common friend recommendation algorithm that can predict whether two users in the network can become friends based on the past network topology. However, traditional link prediction algorithms rely on relatively single indicators

and different indicators have inconsistent prediction capabilities in different networks. The accuracy of the prediction depends on whether the algorithm aligns well with the structural characteristics of the target network. Traditional algorithms have not adequately adapted to network features by selecting appropriate algorithms from the perspective of algorithm combinations or recommending friends for target users through suitable network evolution algorithms. Li, S. et al. [25] proposed a fully integrated homogeneous link prediction algorithm, which focused on solving the issues of sparse networks and the cold-start problem in product marketing. However, it still lacks the application scenario in integrated marketing, and its interpretability is poor, that is, identifying which indicators are effective.

This study has effectively addressed the above shortcomings. For the construction of "integrated marketing social influencers", oriented towards product integration marketing, this study develops the CHLPA from the perspective of heterogeneity of social intimacy to overcome the cold-start problem when predicting links between overlapping nodes of two product circles. Moreover, the algorithm adapts to the network feature indicators to construct composite indicators by suitable indexes for link prediction.

3. Building Social Influencers for Integrated Marketing Based on Friends Recommendations

3.1. Integrated Marketing within a Brand Community

Define online brand community G(V, E) as a graph, where V represents the set of nodes representing users in the community, and E stands for the set of edges representing friendships between users. Users in the brand community form circles based on their shared interests in particular product types. The set of users in the circle of category C products is defined as Vc, and the set of users in the circle of category D products is defined as Vd.

The aim of this study is to achieve efficient integrated marketing by cultivating potential users into integrated marketing social influencers. Specifically, the overlapping node users between the circles of category C and category D products are recommended as friends to other users in both circles. By leveraging the marketing and sales capabilities of social influencers, this approach aims to achieve integrated marketing targeting both the C and D product circles.

Figures 1 and 2 show schematic diagrams of friend recommendations for cultivating integrated marketing social influencers among potential users. Figure 1 displays the network before evolution, while Figure 2 depicts the network after evolution. In Figure 1, the users belonging to the circle of category C products include a, b, c, d, e, f, g, h, and others, while the users belonging to the circle of category D products include a, 1, 2, 3, 4, 5, and 6. Assuming that a is the target user and the goal is to cultivate a social influencer for integrated marketing, friend recommendations aim to make users in both the C and D product circles become friends with user a, thus facilitating integrated marketing between the two product circles. In this study, the SIHI method is used to predict the likelihood of user a becoming friends with users in both the C and D product circles. Users with a high probability of becoming friends are recommended to user a, as shown by the red lines in Figure 2.

3.2. SIHI Aimed at Friend Recommendations

American scholars Nicholas Christakis and James Fowler proposed the "Three Degrees of Influence" principle, which states that only three degrees of separation, or "strong connections," truly affect people in their social networks [26]. Connections beyond three degrees also have an impact, but their influence is weaker and are thus referred to as "weak connections." Strong connections facilitate social behavior, while weak connections transmit information. Based on this principle, this study concludes that the closer a node is to other nodes in terms of social distance, the stronger its social strength with those nodes, making it easier for the node to form social relationships and become friends with other

nodes. This study utilizes the principle of three degrees of influence and a node's closeness centrality to express its level of social intimacy with new friends.
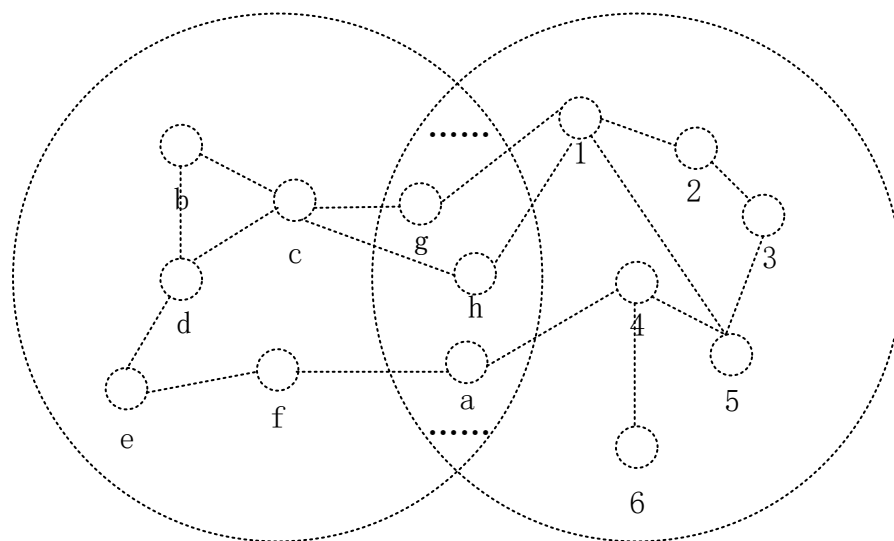

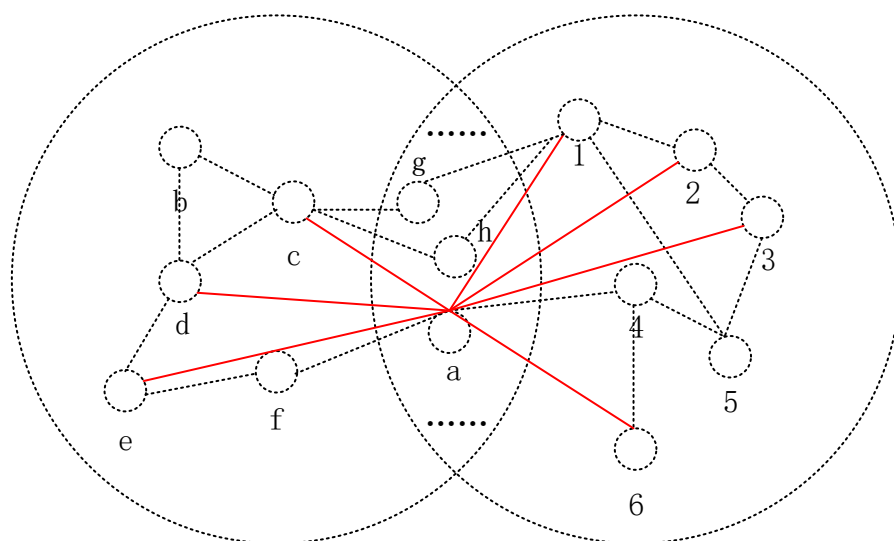**Figure 1.** Network before Evolution.

Figure 2. Network after Evolution.

Variables that describe the importance of nodes mainly include betweenness centrality, degree centrality, and closeness centrality. Among these variables, closeness centrality is closer to the geometric center of the network and can depict global information, making it more suitable for characterizing the intimacy of social connections between nodes. Therefore, this study first used closeness centrality as a global variable to describe the social intimacy of nodes and measure the heterogeneity of the network based on it. If the shortest distance between a node and all other nodes is small, its closeness centrality is high, indicating high social intimacy with other nodes. The closeness centrality of user i is defined as $C(i)$, and the social intimacy assigned by user i to each link in the personal network is $s(i) = \frac{1}{C(i)+1}$. The calculation formula for the closeness centrality of node i is $C(i) = \frac{1}{\sum_{v=1}^{n-1} d(i,v)}$, where n is the total number of nodes in the network, $d(i,v)$ is the shortest distance between node v and node i, which is the shortest path. For any node pairs x and y, and their common neighbor z, SIHI is defined as follows.

3.2.1. SIHI Based on the Node's Own Characteristics

The greater the closeness centrality of a node, the closer its social distance with other nodes, making it easier for the node to become friends with other nodes. Therefore, in real life, two nodes that are close to other nodes in their network are more likely to become friends. Based on this principle, this study creates an SIHI based on the node's own characteristics, as shown in Table 1. M1 represents the sum of two social intimacy values, while M2 represents the product of two social intimacy values.

Table 1. SIHI based on the node's own characteristics.

Algorithm	Formula
M1	$S_{xy}^{M1} = \frac{1}{s(x) + s(y)}$
M2	$S_{xy}^{M2} = \frac{1}{s(x) * s(y)}$

3.2.2. SIHI Based on Nodes and Common Neighbors

The greater the social distance between the common neighbors of a node pair and other nodes, the more social attention is allocated to that node pair, making it more likely for them to become friends [27]. Additionally, the greater the closeness centrality between two nodes, the closer their social distance, making it easier for them to become friends. Based on this principle, this study created a measure of intimacy between nodes and their common neighbors using the SIHI, as shown in Table 2. The M3 calculates similarity between nodes by summing the social intimacy values of their common neighbors and dividing it by their own social intimacy values. M4 calculates similarity by multiplying the social intimacy values of common neighbors and dividing it by the node's own social intimacy values. Similarly, M5 and M6 calculate similarity by summing the social intimacy values of common neighbors in relation to either the maximum or minimum social intimacy values of the nodes.

Table 2. SIHI based on nodes and common neighbors.

Algorithm	Formula
M3	$S_{xy}^{M3} = \frac{\sum_{z \in \Gamma(x) \cap \Gamma(y)} s(z)}{s(x) + s(y)}$
M4	$S_{xy}^{M4} = \frac{\sum_{z \in \Gamma(x) \cap \Gamma(y)} s(z)}{s(x) * s(y)}$
M5	$S_{xy}^{M5} = \frac{\sum_{z \in \Gamma(x) \cap \Gamma(y)} s(z)}{s(x) + s(y)} * \frac{1}{\max\{s(x), s(y)\}}$
M6	$S_{xy}^{M6} = \frac{\sum_{z \in \Gamma(x) \cap \Gamma(y)} s(z)}{s(x) + s(y)} * \frac{1}{\min\{s(x), s(y)\}}$

3.2.3. SIHI Based on Nodes and Community Neighbors

If a node has a smaller clustering coefficient with its neighboring nodes, indicating weaker emotional connection with them, based on the principle of interpersonal interaction, its demand for emotional connection with new friends is higher, and it is more likely to form a link with a new node. Meanwhile, the larger the social distance between the common neighbors of a node pair and other nodes in the network, the more attention the node pair is likely to receive, thereby increasing the probability of a link between them. Additionally, the larger the closeness centrality between two nodes, the closer their social distance, making them more likely to become friends. Based on these principles, this study proposes an SIHI based on nodes and community neighbors to measure the likelihood of establishing a friendship between nodes, as shown in Table 3. M7 considers closeness centrality and social intimacy values for more accuracy. M8 gives more weight to nodes with lower intimacy values. M9 uses both the closeness centrality and intimacy values of common neighbors for a comprehensive similarity measurement. M10 considers the product of intimacy values and centrality for a more comprehensive similarity calculation.

Table 3. SIHI based on nodes and community neighbors.

| Algorithm	Formula
M7	$S_{xy}^{M7} = \dfrac{cc(x) * cc(y)}{\max\{s(x), s(y)\}}$
M8	$S_{xy}^{M8} = \dfrac{cc(x) * cc(y)}{\min\{s(x), s(y)\}}$
M9	$S_{xy}^{M9} = \displaystyle\sum_{z \in \Gamma(x) \cap \Gamma(y)} \left(\dfrac{s(z)}{2 * (s(x) + s(y))} \right) + \dfrac{cc(x)}{s(x)} + \dfrac{cc(y)}{s(y)}$
M10	$S_{xy}^{M10} = \displaystyle\sum_{z \in \Gamma(x) \cap \Gamma(y)} \left(\dfrac{s(z)}{2 * (s(x) * s(y))} \right) + \dfrac{cc(x)}{s(x)} + \dfrac{cc(y)}{s(y)}$

Here, $cc(\cdot)$ represents the inverse of the clustering coefficient of a node, as shown in Formula (1):

$$cc\,(i) = \frac{N_i * (N_i - 1)}{2C} \tag{1}$$

where C is the actual number of links among node i and its neighboring nodes, which is the number of neighboring nodes of node i.

4. CHLPA

To overcome the cold-start problem associated with friend-based recommendations, this study delves deeply into the information contained within the network structure. Specifically, it emphasizes that the social closeness of different nodes varies and proposes CHLPA as an adaptive method for constructing SIHI that is suitable for circle structure characteristics. To accurately capture the compatibility between indexes (such as SIHIs and SLPAs) and the network feature structure, the GBDT model was designed, which employs network feature indicators, such as node clustering and centrality within the local network, to select the appropriate indexes. The process of CHLPA is depicted in Figure 3. CHLPA deals with the cold-start issue in network analysis by utilizing heterogeneous indicators, combining them through weighted or comprehensive methods, selecting appropriate indicators based on network characteristics, and leveraging the power of the GBDT for accurate predictions and decision-making. This approach improves the understanding and utilization of network information, particularly in the context of integrated marketing and brand community analysis.
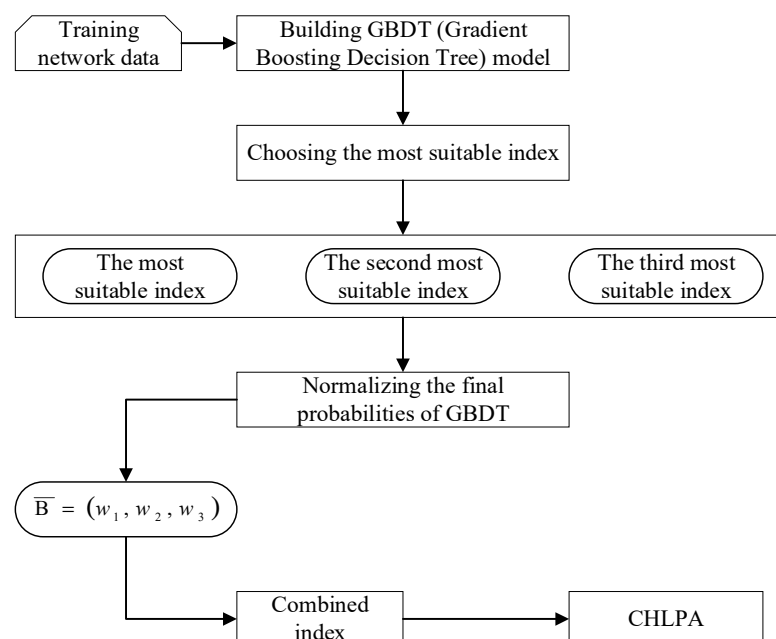
Figure 3. CHLPA process flow.

4.1. Network Features for Filtering SIHI

4.1.1. Node Density

(a) Average Path Length

The average path length represents the average distance between all pairs of nodes and is used to measure the dispersion of emotional intimacy between each node and other nodes in the network. The formula for the average path length is shown in Formula (2).

$$L = \frac{2\sum_{i \geq j} d_j}{N * (N-1)} \tag{2}$$

The variable d_j represents the path length between nodes, and N is the number of nodes.

(b) Network Diameter

The network diameter refers to the maximum distance between any two nodes in the network. The Formula (3) shows the calculation of the network diameter.

$$d(i,j) = \max_{i \neq j} L_{ij} \tag{3}$$

where $L_{ij} = \min_{i \neq j} l_{ij}$ represents the shortest distance between any two nodes, where l_{ij} is the length of all paths from node i to node j.

4.1.2. Node Centrality

(a) Average Node Degree Centrality

Degree centrality is the most direct measure of node centrality in network analysis. The larger the degree of a node, the higher its degree centrality and the more important it is in the network. Formula (4) shows the calculation of average node degree centrality.

$$\bar{s} = \frac{1}{N} \sum_{i=1}^{N} s_i \tag{4}$$

where N is the number of nodes, and s_i represents the degree centrality of node i.

(b) Average node betweenness centrality

Betweenness centrality refers to the number of times a node serves as a bridge along the shortest path between two other nodes. The betweenness centrality of a node is the number of these shortest paths that pass through that node. The more times a node acts as an "intermediary," the greater its betweenness centrality. Betweenness centrality is a commonly used measure for characterizing the importance of nodes. The formula for the average node betweenness centrality is shown in Formula (5).

$$C_G = \frac{1}{|V|} * \sum_{i=1}^{|V|} \frac{2}{(|V|-1)*(|V|-2)} * \sum_{j<k} \frac{n_{jk}(i)}{n_{jk}} \tag{5}$$

where V represents the number of edges, $n_{jk}(i)$ stands for the number of shortest paths from node j to node k that pass-through node i, and n_{jk} represents the total number of shortest paths from node j to node k.

(c) Average Eigenvector Centrality

The fundamental concept behind eigenvector centrality is that the centrality of a node is a function of the centralities of its neighboring nodes. In other words, the more important the other nodes connected to a node, the more important the node itself is. Therefore, eigenvector centrality can effectively measure the importance of nodes while considering the importance of their neighbors and can be used to measure the transmission impact

and connectivity between nodes. The formula for the average eigenvector centrality is displayed in Formula (6).

$$EC(i) = x_i = c \sum_{j=1}^{N} a_{ij} g_j \tag{6}$$

where c is a proportionality constant, $a_{ij} = 1$ if there is a connection between node i and node j, otherwise $a_{ij} = 0$, g_j is the degree of node j, and N is the total number of nodes.

4.2. GBDT

The GBDT model was proposed by Jerome Friedman as an iterative decision tree algorithm composed mainly of multiple Classification and Regression Trees (CART) [28]. During the prediction phase, the model traverses all nodes of a tree for each input feature vector, and each tree generates a prediction value based on the input vector's features. The final prediction value is obtained by aggregating all the prediction values from the trees.

The GBDT offers various advantages in certain aspects. Firstly, GBDT performs well in processing nonlinear data and can recognize the interaction between features. Secondly, GBDT can achieve high accuracy and generalization performance in training and testing sets. Thirdly, GBDT is constructed using decision trees, making the results easy to understand and highly explanatory. The GBDT can be used to solve both regression and classification problems. One of the fundamental concepts of The GBDT is to use the negative gradient values of the loss function as an estimated value of the residuals to fit a classification tree. When decision tree algorithms are used alone, they are prone to overfitting. GBDT decreases the complexity of decision trees, reduces the fitting ability of a single decision tree, and then integrates multiple decision trees through gradient boosting, ultimately solving the problem of overfitting very well [29].

In the GBDT, the independent variables used for learning the samples are network feature indicators, while the dependent variable is the one-hot encoding of the π categories of SIHIs and SLPAs (i.e., the encoding corresponding to the index with the maximum AUC value is 1, and the others are 0). The π categories indexed in this study can each generate a CART for GBDT. The core idea of CART is that each round of training is based on the residual of the previous round of training. Here, the residual refers to the negative gradient value of the current model. For the GBDT's multiclass classification problem, the softmax model needs to be considered as follows:

$$P(y = j | x) = \frac{e^{F_j(x)}}{\sum_{i=1}^{k} e^{F_i(x)}} \tag{7}$$

where $F_j(x)$ represents the predicted value of the jth CART for category x. As this study adapts π indexes, the value of K is π. The single-sample loss function of the softmax model is as follows:

$$loss = - \sum_{i=1}^{k} y_i \log \frac{e^{F_k(x)}}{\sum_{i=1}^{k} e^{F_i(x)}} \tag{8}$$

where $y_i (i = 1, 2, \ldots, k)$ stands for the value of class i in one-hot encoding of the SIHIs.

The derivative of the loss function with respect to F_i is as follows:

$$-\frac{\partial loss}{\partial F_i} = y_i - P(y_i | x) \tag{9}$$

In summary, the basic steps for training the GBDT model in this study are as follows:

Step 1: We select an index based on five different network feature indicators, including average path length, network diameter, average node degree centrality, average node betweenness centrality, and average eigenvector centrality for a comprehensive evaluation of image dehazing algorithms. Therefore, we use a five-dimensional vector to represent the input of each sample and select the most suitable indicator from π indexes based on the

five different indicators. For each index, we independently trained a CART with a depth of five to avoid overfitting. The root and internal nodes of the tree correspond to the five different network feature indicators, and the leaf nodes represent the clustering labels of the samples. We set the number of trees generated through training for each index to 90, ensuring a robust selection process, named CART Tree1, CART Tree2, ... , CART Tree90. The first most suitable index is the one selected the most times, the second most suitable index is the one selected the second most times, and the third most suitable index is the one selected the third most times, providing a quantitative measure of the relative performance. The variable m represents the number of iterations, and we initialize m to 1 to find more effective algorithms.

Step 2: Train CART. First, initialize the data. The input for the training samples consists of five network feature indicators, and the output is the index independent hot encoding (defined previously), which allows for a comprehensive evaluation of the index algorithm performance. During CART training, use Formula (10) to calculate the probability that a training sample belongs to the index. Due to being a classification problem, it is not possible to compare the size of each classification problem; therefore, the probability of which category to classify is used to represent the classification results.

$$P(y = k|x_{1...}x_5) = \frac{e^{F_{k,m}(x_{1...}x_5)}}{\sum_{o=1}^{10} e^{F_{o,m}(x_{1...}x_5)}} \tag{10}$$

The formula $F_{k,m}(x_{1...}x_5) = \frac{G_k(x_{1...}x_5)}{Q}$ represents the predicted value of the k-th index in the mth round of training. Here, $G_k(x_{1...}x_5)$ is the count of the one-hot encoding of the kth index with a value of one among all samples, and $x_{1...}x_5$ are the five network feature indicators. Q is the total nber of samples.

Step 3: Calculate the negative gradient values of the index using Formula (11). Negative gradient values are a technique used to minimize loss in gradient descent optimization algorithms and improve model performance and accuracy.

$$\widetilde{y_{i,k}} = y_{ik} - P(y_i = k|x_{i1...}x_{i5}) \; i = 1, \cdots, n, \; k = 1, \cdots, K \tag{11}$$

where y_{ik} refers to the one-hot encoded value of the kth class index in sample i, $P(y_i = k|x_{1...}x_5)$ is the probability that sample i belongs to the kth class index, and $\widetilde{y_{i,k}}$ is the negative gradient value of the kth class index in sample i. n represents the number of samples, x_{ij} represents the jth input in sample i, and K = 10.

Step 4: Generate a CART for each index, resulting in K CARTs. The procedure for generating each CART is as follows. Use the five types of network feature indicators as the nodes of the CART, choose one network feature indicator from the five types of feature indicators, and traverse all possible values to find a critical value. Sample classification is then performed based on this critical value. The criterion for classification is to minimize the value of Formula (12). The critical value is used as the splitting condition, where samples with values less than the critical value are split to the left leaf node, and samples with values greater than the critical value are split to the right leaf node.

$$min\left(min \sum_{x_i \in H_i} (\widetilde{y_{k,m}} - c_1)^2 + min \sum_{x_i \in H_i} (\widetilde{y_{k,m}} - c_2)^2 \right) \tag{12}$$

where c_1 represents the mean negative gradient value of the samples allocated to the left leaf node, while c_2 represents the mean negative gradient value of the samples allocated to the right leaf node.

This process is repeated until all five feature indicators are considered as splitting conditions, completing one round of CART training. K CARTs can be generated by following the same steps.

Step 5: Update $F_{k,m}(x1\ldots x5)$ [28] based on the generated CART using the following formulas.

$$\gamma_{jkm} = \frac{K-1}{K} \frac{\sum_{s_i \in R_{jkm}} \widetilde{y_{ik}}}{\sum_{s_i \in R_{jkm}} |\widetilde{y_{ik}}|(1-|\widetilde{y_{ik}}|)}, j = 1\ldots, J, k = 1,\ldots,K \tag{13}$$

$$F_{k,m+1}(x_1\ldots x_5) = F_{k,m}(x_1\ldots x_5) + \sum_{j=1}^{J} \gamma_{jkm} I\left(x_1\ldots x_5 \in R_{jkm}\right), k = 1,\ldots,K \tag{14}$$

$$P(y = k|x_{1\ldots}x_5) = \frac{e^{F_{k,m+1}(x_{1\ldots}x_5)}}{\sum_{o=1}^{K} e^{F_{o,m+1}(x_{1\ldots}x_5)}}, \ k = 1,\ldots,K \tag{15}$$

where R_{jkm} represents the set of samples corresponding to the jth leaf node of the CART for the kth type of index in the mth round of training. $I\left(x_1\ldots x_5 \in R_{jk1}\right)$ is an indicator function that equals one if the sample containing $x_1\ldots x_5$ belongs to Rjkm and 0 otherwise. s_i represents the ith sample, and J is the number of bottom-level leaf nodes in the CART.

Step 6: Begin the next round of training by setting m = m + 1 and repeating steps 3–6. Iterate in this manner until the specified time limit is reached (e.g., M rounds), where M is set to 90 in this study. By continuing the training process through multiple rounds, the model is able to learn and improve its accuracy over time.

Step 7: The probability that a new sample is suitable for a certain index is shown in Formula (16). In this study, the index with the highest probability is selected as the most suitable index.

$$p_k(x_{1\ldots}x_5) = \frac{e^{F_{k,M}(x_{1\ldots}x_5)}}{\sum_{o=1}^{10} e^{F_{o,M}(x_{1\ldots}x_5)}} \tag{16}$$

Step 8: Among all indexes, the indexes with the highest, second-highest, and third-highest probabilities calculated according to formula (16) are selected as the first suitable, second suitable, and third suitable indexes for the brand community, i.e., $E_{first\ suitable}$, $E_{second\ suitable}$, and $E_{third\ suitable}$.

The final combination of indexes is as follows:

$$MAA = w_1 * E_{first\ suitable} + w_2 * E_{second\ suitable} + w_3 * E_{third\ suitable} \tag{17}$$

Here, w_1, w_2, and w_3 are the weights of the first, second, and third suitable indexes, which are determined according to the fitting probabilities calculated from Formula (16).

Similarly, link prediction is made based on the combined index of Formula (17), and then a social influencer for integrated marketing is constructed by evolving the network based on changes in the closeness centrality index using the hill-climbing algorithm.

The pseudo code for CHLPA is shown in Algorithm 1.

Algorithm 1. The proposed CHLPA algorithm.

Input: a set of no-direction network
Output: a set of combined indexes for node pairs
Function CHLPA
For node pairs (x,y) in **non-edges**
    Node Density
    Node Centrality
    GBDT choosing the suitable indexes
Return a set of combined indexes

---

5. Experimental Design and Result Analysis

5.1. Experimental Design

To verify the effectiveness of CHLPA, one hundred and five ego net datasets from Google Plus (http://snap.stanford.edu/data/ego-Gplus.html (accessed on 10 January 2023), one hundred ego net datasets from Twitter (http://snap.stanford.edu/data/ego-Twitter.htm (accessed on 10 January 2023), and one ego net datasets Celegans(https://deim.urv.cat/~alexandre.arenas/data/welcome.htm (accessed on 10 January 2023) provided by Stanford University were selected as the experimental objects. In each experiment, 90% of the datasets were randomly selected from the network as the training set, and the remaining 10% were used as the test set. The experiment was repeated 100 times to obtain the average value. Matlab 2022 software was used to implement the experiment, and the algorithm parameters, except as specifically noted, were selected using the default values provided by the software. Table 4 shows the average, minimum, and maximum values of the selected network samples from Google Plus, Twitter, and Celegans.

Table 4. Statistics of the Network Sample.

Dataset	Statistical Indicators	MIN	AVERAGE	MAX
Google Plus	average path length	1.4503	2.0078	2.5968
	network diameter	2.000	4.4583	7.000
	mean node degree centrality	5.824	23.2574	56.557
	average node betweenness centrality	8.105	258.9533	1054.731
	mean eigenvector centrality	0.856	2.3905	4.720
Twitter	average path length	1.1209	1.9739	2.8883
	network diameter	2.0000	4.4810	8.0000
	mean node degree centrality	2.8889	24.6438	92.4426
	average node betweenness centrality	1.5714	118.6830	378.5911
	mean eigenvector centrality	0.8953	2.7602	6.7377
Celegans	average path length		2.7375	
	network diameter		8.0000	
	mean node degree centrality		27.3552	
	average node betweenness centrality		2121.5237	
	mean eigenvector centrality		1.2455	

AUC is the primary metric to evaluate algorithm performance accuracy; it measures the randomness of pairs of linked and unlinked nodes in the test set compared to their scores generated by SIHI or SLPA [27]. Specifically, the AUC is calculated by randomly sampling N pairs of nodes n from the test set and comparing their scores. If the score for the linked node is higher than the score for the unlinked node N1 times out of the N comparisons, then the AUC value is calculated using Formula (18):

$$AUC = \frac{N1 + 0.5(N - N1)}{N} \tag{18}$$

5.2. Algorithm Performance Analysis

In this section, the experimental results of the proposed method are presented and compared with benchmark methods and different datasets. The results indicate that the CHLPA performs better than the benchmark methods and can effectively address the issue of cold-start problems in nodes. The formulas of benchmark algorithms are as shown in Table 5.

Table 6 and Figure 4 provide a comparison of the performances of the different algorithms on different network datasets. Figure 4 is the candlestick of the AUC results for CHLPA considering the average, the maximum, the minimum, and the standard deviation of the AUC values. The results presented in Table 6 and Figure 4 show that the performance of MAA outperforms any individual SIHI, indicating that the proposed method

of generating a combined SIHI using GBDT is effective and that CHLPA significantly improves the algorithm's performance. In other words, when constructing social influencers for integrated marketing in social networks, CHLPA accurately predicts the likelihood that overlapping node target users will establish friendships with other node users in the community. Table 7 provides a performance comparison between the traditional SLPA and MAA0 on Google Plus datasets, a combination algorithm constructed based on the traditional SLPA, and the proposed combined index architecture in this study. From Figure 4 and Table 7, it can be seen that MAA0's performance is significantly better than that of SLPA, indicating that the proposed architecture for constructing combined indexes, which combines five network characteristics, GBDT, and three types of suitability indexes, effectively improves the accuracy of link prediction for overlapping nodes. Additionally, the performance of MAA is significantly better than that of MAA0, indicating that using global variables such as closeness centrality to characterize the network position of nodes, proposing SIHI based on the three-degree influence principle, and generating suitable combined indexes through GBDT accurately and effectively characterize the implicit social information in overlapping networks and improve algorithm performance.

Table 5. The formulas of benchmark algorithms.

Algorithm	Formula	Reference
CN	$S_{xy} = \|\Gamma(x) \bigcap \Gamma(y)\|$	[30]
Salton	$S_{xy} = \frac{\|\Gamma(x) \bigcap \Gamma(y)\|}{\sqrt{\|\Gamma(x) * \Gamma(y)\|}}$	[31]
Jaccard	$S_{xy} = \frac{\|\Gamma(x) \bigcap \Gamma(y)\|}{\|\Gamma(x) \bigcup \Gamma(y)\|}$	[32]
Sorenson	$S_{xy} = \frac{2\|\Gamma(x) \bigcap \Gamma(y)\|}{k_x + k_y}$	[30]
Hub Promoted Index (HPI)	$S_{xy} = \frac{\|\Gamma(x) \bigcap \Gamma(y)\|}{\min(\Gamma(x), \Gamma(y))}$	[33]
Hub Depressed Index (HDI)	$S_{xy} = \frac{\|\Gamma(x) \bigcap \Gamma(y)\|}{\max(\Gamma(x), \Gamma(y))}$	[33]
LHN	$S_{xy} = \frac{\|\Gamma(x) \bigcap \Gamma(y)\|}{\|\Gamma(x)\| * \|\Gamma(y)\|}$	[34]
Adamic/Adar (AA)	$S_{xy} = \sum\limits_{z \in \Gamma(x) \bigcap \Gamma(y)} \frac{1}{log\|\Gamma(z)\|}$	[35]
Resource Allocation (RA)	$S_{xy} = \sum\limits_{z \in \Gamma(x) \bigcap \Gamma(y)} \frac{1}{\|\Gamma(z)\|}$	[36]
DGLP	$S_{xy} = \frac{\|\Gamma(x) + \Gamma(y)\|}{d_{xy} + 1} + \sum\limits_{z \in \Gamma(x) \bigcap \Gamma(y)} \Gamma(z)$	[37]
CCPA	$S_{xy} = \alpha * (\|\Gamma(x) \bigcap \Gamma(y)\|) + (1 - \alpha) * \frac{N}{d_{xy}}$	[38]

Note: For CCPA, parameter $\alpha \in [0, 1]$ is the defined parameter value to control the weight of centrality and common neighbor, N is the number of nodes, and d_{xy} is the shortest distance between node x and y. *For Sorenson, k_x* represents the degree of a node x.

Table 6. AUC of all SIHI in different datasets.

Algorithm	Google Plus	Celegans	Twitter
M1	0.8057	0.7263	0.8321
M2	0.8059	0.7265	0.8316
M3	0.6710	0.7056	0.7453
M4	0.6719	0.7066	0.7460
M5	0.6708	0.7067	0.7455
M6	0.6734	0.7062	0.7443
M7	0.6317	0.6035	0.7710
M8	0.6338	0.6026	0.7719
M9	0.5017	0.2696	0.6048
M10	0.6656	0.7062	0.7496
MAA	0.8515	0.8839	0.8868
DGLP	0.6895	0.6828	0.7100
CCPA	0.7231	0.5328	0.8241

Note: MAA is the combined SIHI proposed in this study. Google plus' MAA = 0.999 * RA + 0.0005 * M1 + 0.0005 * AA. Celegans' MAA = 0.999 * RA + 0.0005 * AA + 0.0005 * M1. Twitter's MAA = 0.999 * RA + 0.0005 * AA + 0.0005 * M1. For CCPA, we consider α = {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}.
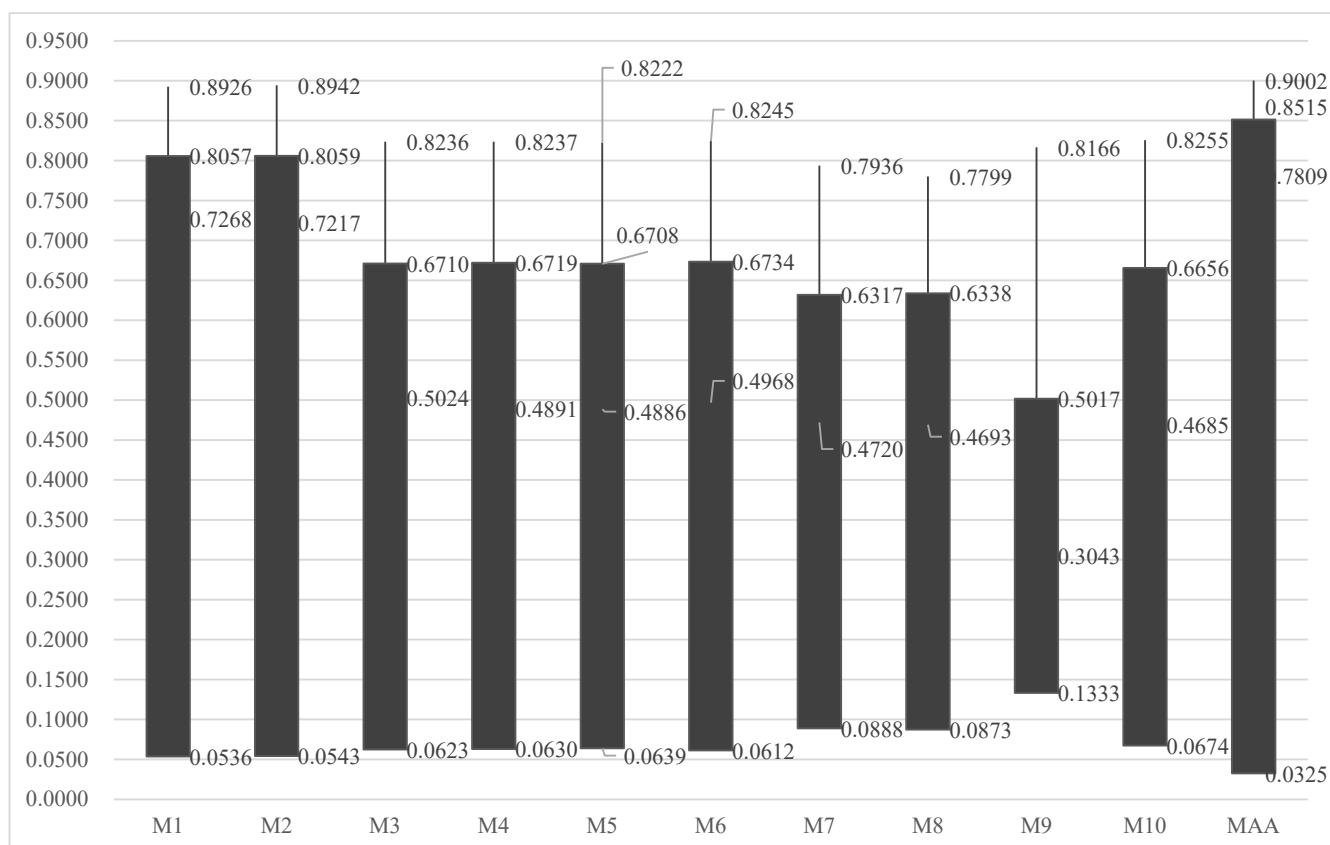
Figure 4. Candlestick of AUC results for CHLPA.

Table 7. Performance Comparison of CHLPA proposed in this study and all reference methods.

Algorithm	Average AUC	Algorithm	Average AUC
CN	0.7987	HDI	0.6330
Salton	0.6599	LHN	0.4562
Jaccard	0.6480	AA	0.8081
Sorenson	0.6489	RA	0.8123
HPI	0.6956	DGLP	0.6895
CCPA	0.7231	MAA0	0.8174

5.3. Establishing Friend Relationships Based on Hill-Climbing Algorithm

According to the different network densities, extremely sparse networks (with ID 114124942936679476879), sparse networks (with ID 104917160754181459072), and dense networks (with ID 112573107772208475213) on Google Plus datasets were selected as examples to study the effectiveness of the hill-climbing method proposed in this study for constructing friendships in evolving networks. For each of the three types of networks, we chose a target user as the overlapping node and analyzed the evolution process of the network based on the proposed CHLPA.

Table 8 shows the evolution of the target user's friends during the hill-climbing algorithm based on node degree, node degree centrality, node betweenness centrality, and node closeness centrality. C_i represents the number of friends for the ith iteration, and the stopping iteration is based on the termination round of the hill-climbing algorithm based on node closeness centrality. For the extremely sparse network (ID 114124942936679476879), the threshold for the number of friends of the target user was set to 30, for the sparse network (ID 104917160754181459072), the threshold of the number of friends of the target user was set to 110, and for the dense network (ID 112573107772208475213), the threshold for the number of friends of the target user was set to 190. From Table 8, it is evident

that the hill-climbing algorithm based on node closeness centrality has a faster and more efficient friend recommendation performance than the other hill-climbing algorithms.

Table 8. Prediction of Friend Relationships Constructed using Hill-Climbing Algorithm.

Network Code	Number of Nodes	Number of Circles	Target User ID	Algorithm	Number of Friends after Each Iteration					
					C1	C2	C3	C4	C5	C6
114124942936 679476879	34	2	101889975 950769	Hill-Climbing Algorithm Based on Node Degree	1	11	23			
				Hill-Climbing Algorithm Based on Node Degree Centrality	1	18	25			
				Hill-Climbing Algorithm Based on Node Betweenness Centrality	1	17	27			
				Hill-Climbing Algorithm Based on Node Closeness Centrality	1	21	33			
104917160754 181459072	132	6	111043623 176980	Hill-Climbing Algorithm Based on Node Degree	1	16	29	48	79	103
				Hill-Climbing Algorithm Based on Node Degree Centrality	1	16	35	57	79	102
				Hill-Climbing Algorithm Based on Node Betweenness Centrality	1	20	36	59	84	104
				Hill-Climbing Algorithm Based on Node Closeness Centrality	1	34	56	87	103	116
112573107772 208475213	202	14	115716197 313320	Hill-Climbing Algorithm Based on Node Degree	1	48	79	102	148	179
				Hill-Climbing Algorithm Based on Node Degree Centrality	1	35	667	99	105	132
				Hill-Climbing Algorithm Based on Node Betweenness Centrality	1	45	79	106	124	158
				Hill-Climbing Algorithm Based on Node Closeness Centrality	1	67	89	102	142	198

Upon comparison, it can be observed that the climbing algorithm based on closeness centrality facilitates a faster rate of friend recommendations between the target user and ordinary users. Therefore, in the context of integrated marketing, applying CHLPA and the climbing algorithm proposed in this study to recommend friends for overlapping network users can swiftly convert potential ordinary users into social influencers for integrated marketing. This approach can then be used to market products by leveraging the social influence of integrated marketing social influencers, thereby saving marketing costs.

6. Conclusions

In the early stages of brand establishment, the customer base is usually limited. Therefore, product integration marketing can quickly attract fans to the brand. To construct "integrated marketing social influencers" and accumulate a large fan base for overlapping

nodes between two product circles within a short period of time, it is necessary to identify node users in both product circles who can establish friendships with overlapping node users. Therefore, this study proposes CHLPA to predict possible friendship relationships between overlapping node users and other node users in different product circles, thereby achieving efficient and accurate product integration marketing.

To address the cold-start problem when predicting friend relationships between target overlapping node users and other node users in two product communities, this study takes an adaptive approach based on network characteristics. This study proposes the SIHI method, which utilizes closeness centrality as a global variable to represent the social intimacy between nodes, and develops the CHLPA model. Following the three-degree influence principle, CHLPA proposes three types of SIHI, which are based on node-based, neighbor-based, and community-based approaches, to comprehensively describe the probability of users becoming friends in a sparse network environment. Then, CHLPA utilizes the GBDT method to adaptively select the most suitable index, the second most suitable index, and the third most suitable index from SIHIs and SLPAs based on the density of nodes and their centrality in the local network. CHLPA then assigns weights to the selected index according to the normalized final fit probabilities from GBDT and combines them to predict the probability of overlapping node users and other node users becoming friends in two product communities in an online brand community. As a result, an integrated marketing social influencer can be built efficiently.

To evaluate the performance of CHLPA, this study conducted simulation experiments using the Google Plus, Twitter, and Celegans datasets. The experimental results show that CHLPA has excellent performance. In addition, this study investigated the effectiveness of the hill-climbing method in constructing friendship relationships in evolving networks, focusing on extremely sparse, sparse, and dense networks. One overlapping node target user was selected based on the proposed CHLPA. The results show that the node-closeness centrality-based hill-climbing algorithm has a faster and more efficient friend recommendation efficiency than the others. Therefore, applying the proposed CHLPA and hill-climbing algorithms to friend recommendations for overlapping nodes in social networks can efficiently transform potential ordinary users into integrated marketing social influencer users, who can influence marketing through the friends of integrated marketing social influencers, and thus reduce marketing costs.

Further research in the field of link prediction is expected to explore the integration of new data sources and features to enhance accuracy. One promising approach is the incorporation of social context information, such as user demographics and online behavior, alongside link structure data. This can provide a more comprehensive understanding of user behavior and improve link prediction performance. Moreover, the development of more advanced machine learning techniques should also be considered to improve the efficiency and accuracy of link prediction models. Recent innovations in deep learning and reinforcement learning have shown great potential in enhancing the performance of link prediction algorithms and may be utilized in future research for even better outcomes.

Author Contributions: Conceptualization, S.L., H.Z., J.Z. and Z.Y.; Data curation, Y.W., Y.Z. and Z.W.; Formal analysis, J.Z. and Z.Y.; Funding acquisition, S.L.; Investigation, J.L.; Methodology, H.Z. and Y.Z.; Project administration, S.L.; Resources, J.L.; Software, Z.W.; Supervision, S.L.; Validation, S.L. and H.Z.; Visualization, J.Z. and Z.Y.; Writing—original draft, H.Z.; Writing—review & editing, J.Z., Z.Y. and Y.W. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available upon reasonable request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, D.H.; Brusilovsky, P. How to Measure Information Similarity in Online Social Networks: A Case Study of Citeulike. *Inf. Sci.* **2017**, *418*, 46–60. [CrossRef]
2. Muniz, A.M., Jr.; O'Guinn, T.C. Brand Community. *J. Consum. Res.* **2001**, *27*, 412–432. [CrossRef]
3. Horne, B.D.; Nevo, D.; Adalı, S. Recognizing Experts on Social Media: A Heuristics-Based Approach. *ACM SIGMIS Database* **2019**, *50*, 66–84. [CrossRef]
4. Bakshy, E.; Hofman, J.M.; Mason, W.A.; Watts, D.J. Everyone's an Influencer: Quantifying Influence on Twitter. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, Hong Kong, China, 9–12 February 2011; pp. 65–74.
5. Palazón, M.; Sicilia, M.; López, M. The Influence of "Facebook Friends" on the Intention to Join Brand Pages. *J. Prod. Brand Manag.* **2015**, *24*, 563–577. [CrossRef]
6. Choi, Y.K.; Seo, Y.; Yoon, S. E-WOM Messaging on Social Media: Social Ties, Temporal Distance, and Message Concreteness. *Internet Res.* **2017**, *27*, 516–532. [CrossRef]
7. Renton, M.; Simmonds, H. Like is a Verb: Exploring Tie Strength and Casual Brand Use Effects on Brand Attitudes and Consumer Online Goal Achievement. *J. Prod. Brand Manag.* **2017**, *26*, 610–621. [CrossRef]
8. Whittler, T.E.; Spira, J.S. Model's Race: A Peripheral Cue in Advertising Messages? *J. Consum. Psychol.* **2002**, *12*, 291–301. [CrossRef]
9. Voyer, P.A.; Ranaweera, C. The Impact of Word of Mouth on Service Purchase Decisions: Examining Risk and the Interaction of Tie Strength and Involvement. *J. Serv. Theory Pract.* **2015**, *25*, 274–293.
10. Lee, J.N. A Study on the Impact of Brand Community Interaction Model on Brand Loyalty: Focusing on Chinese Online Brand Community. *Korean Corp. Manag. Rev.* **2012**, *46*, 93–113.
11. Wilson, H.N.; Macdonald, E.K.; Baxendale, S. What Really Makes Customers Buy a Product. *Harv. Bus. Rev.* **2015**, *93*, 1–11.
12. Ghalmane, Z.; Cherifi, C.; Cherifi, H.; El Hassouni, M. Exploring Hubs and Overlapping Nodes Interactions in Modular Complex Networks. *IEEE Access* **2020**, *8*, 79650–79683. [CrossRef]
13. Lü, L.; Zhou, T. Link Prediction in Complex Networks: A Survey. *Phys. A Stat. Mech. Its Appl.* **2011**, *390*, 1150–1170. [CrossRef]
14. Sun, Y.; Barber, R.; Gupta, M.; Aggarwal, C.C.; Han, J. Co-Author Relationship Prediction in Heterogeneous Bibliographic Networks. In Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining, Kaohsiung, Taiwan, 25–27 July 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 121–128.
15. He, Y.; Liu, J.N.K.; Hu, Y.; Kao, B.; Lee, W.-C.; Chen, A.L.P.; Chen, M.-S. OWA Operator-Based Link Prediction Ensemble for Social Network. *Expert Syst. Appl.* **2015**, *42*, 21–50. [CrossRef]
16. Wang, J.; Ma, Y.; Liu, M.; Tang, X.; Qian, C. Link Prediction Based on Community Information and Its Parallelization. *IEEE Access* **2019**, *7*, 62633–62645. [CrossRef]
17. Xiao, Y.; Li, X.; Wang, H.; Liu, Y.; Liu, Y. 3-HBP: A Three-Level Hidden Bayesian Link Prediction Model in Social Networks. *IEEE Trans. Comput. Soc. Syst.* **2018**, *5*, 430–443. [CrossRef]
18. Bütün, E.; Kaya, M. A Pattern-Based Supervised Link Prediction in Directed Complex Networks. *Phys. A Stat. Mech. Its Appl.* **2019**, *525*, 1136–1145. [CrossRef]
19. Zhan, Q.; Zhang, J.; Philip, S.Y.; Jin, Q.; Yang, Y.; Yu, Y. Discover Tipping Users for Cross Network Influencing. In Proceedings of the 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI), Pittsburgh, PA, USA, 28–30 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 67–76.
20. Dewi, F.K.; Yudhoatmojo, S.B.; Budi, I. Identification of Opinion Leader on Rumor Spreading in Online Social Network Twitter Using Edge Weighting and Centrality Measure Weighting. In Proceedings of the 2017 Twelfth International Conference on Digital Information Management (ICDIM), Fukuoka, Japan, 12–14 September 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 313–318.
21. Dai, W.; Shang, Y. Edge-Concerned Embedding for Multiplex Heterogeneous Network. In Proceedings of the 2020 IEEE 6th International Conference on Computer and Communications (ICCC), Chengdu, China, 17–20 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 584–589.
22. Shi, S.; Li, Y.; Wen, Y.; Li, J.; Gao, J. Adding the Sentiment Attribute of Nodes to Improve Link Prediction in Social Networks. In Proceedings of the 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Zhangjiajie, China, 15–17 August 2015; pp. 1263–1269.
23. Mohdeb, D.; Boubetra, A.; Charikhi, M. WMPLP: A Model for Link Prediction in Heterogeneous Social Networks. In Proceedings of the 2014 4th International Symposium ISKO-Maghreb: Concepts and Tools for Knowledge Management (ISKO-Maghreb), Tetouan, Morocco, 22–24 October 2014; pp. 1–4.
24. Cao, B.; Kong, X.; Philip, S.Y. Collective Prediction of Multiple Types of Links in Heterogeneous Information Networks. In Proceedings of the 2014 IEEE International Conference on Data Mining, Shenzhen, China, 14–17 December 2014; pp. 50–59.
25. Li, S.; Wang, Z.; Zhang, B.; Zhu, B.; Wen, Z.; Yu, Z. The Research of "Products Rapidly Attracting Users" Based on the Fully Integrated Link Prediction Algorithm. *Mathematics* **2022**, *10*, 2424. [CrossRef]
26. Wang, H.; Yin, G.; Zhou, L.; Shao, Z. Influence Maximization Algorithm in Social Networks Based on Three Degrees of Influence Rule. In Proceedings of the International Conference on Cloud Computing and Security, Nanjing, China, 8–10 June 2018; Springer: Cham, Switzerland, 2018; pp. 567–578.
27. Li, S.; Song, X.; Lu, H.; Huang, J.; Zhang, H. Friend Recommendation for Cross Marketing in Online Brand Community Based on Intelligent Attention Allocation Link Prediction Algorithm. *Expert Syst. Appl.* **2020**, *139*, 112839. [CrossRef]

28. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

29. Zhang, C.; Liu, C.; Zhang, X.; He, Q.; Fan, Y. An Up-to-Date Comparison of State-of-the-Art Classification Algorithms. *Expert Syst. Appl.* **2017**, *82*, 128–150. [CrossRef]

30. Lü, L.; Jin, C.-H.; Zhou, T. Similarity index based on local paths for link prediction of complex networks. *Phys. Rev. E* **2009**, *80*, 046122. [CrossRef]

31. Salton, G.; McGill, M.J. *Introduction to Modern Information Retrieval*; McGraw-Hill: New York, NY, USA, 1983.

32. Jaccard, P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull. Soc. Vaudoise Sci. Nat.* **1901**, *37*, 547–579.

33. Ravasz, E.; Somera, A.L.; Mongru, D.A.; Oltvai, Z.N.; Barabási, A.L. Hierarchical organization of modularity in metabolic networks. *Science* **2002**, *297*, 1551–1555. [CrossRef] [PubMed]

34. Leicht, E.A.; Holme, P.; Newman, M.E.J. Vertex similarity in networks. *Phys. Rev. E* **2006**, *73*, 026120. [CrossRef] [PubMed]

35. Adamic, L.A.; Adar, E. Friends and neighbors on the web. *Soc. Netw.* **2003**, *25*, 211–230. [CrossRef]

36. Zhou, T.; Lü, L.; Zhang, Y.-C. Predicting missing links via local information. *Eur. Phys. J. B* **2009**, *71*, 623–630. [CrossRef]

37. Yuliansyah, H.; Othman, Z.A.; Bakar, A.A. A new link prediction method to alleviate the cold-start problem based on extending common neighbor and degree centrality. *Phys. A Stat. Mech. Its Appl.* **2023**, *616*, 128546. [CrossRef]

38. Ahmad, I.; Akhtar, M.U.; Noor, S.; Shahnaz, A. Missing link prediction using common neighbor and centrality based parameterized algorithm. *Sci. Rep.* **2020**, *10*, 364. [CrossRef]